DOCUMENT RESUME

ED 263 641 CS 505 120

AUTHOR Roberts, Charles V.
TITLE A User's Response to the Use of Listening Assessment Instruments.
PUB DATE Nov 85
NOTE 18p.; Paper presented at the Annual Meeting of the Speech Communication Association (71st, Denver, CO, November 7-10, 1985).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Listening Comprehension Tests; *Listening Skills; Research Problems; Student Evaluation; *Test Construction; *Testing; Test Theory; Test Use; Test Validity

ABSTRACT

Noting that the attention of the speech communication discipline to listening skills does not mirror the apparent importance of such skills, this paper examines five listening assessment tests--focusing on the strengths, weaknesses, procedural problems, and conceptualizations of each--that potential users should be aware of before selecting any one of the instruments. The major portion of the paper discusses individually the five instruments: (1) the Learning Skills Inventory, (2) the Communication Competency Assessment Instrument, (3) the Brown-Carlsen Listening Comprehension Test, (4) the Watson-Barker Listening Test, and (5) the Kentucky Comprehensive Listening Test. The remainder of the paper presents conclusions drawn from the previous discussions, specifically that one instrument cannot win acceptance without its underlying conceptual definition being agreed to by the majority of users, and that perhaps no one instrument will be found to be acceptable for all situations. This section also acknowledges many of the shortcomings of existing listening research, and the need for longitudinal investigations that would document effective methods for teaching listening. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED263641

A USER'S RESPONSE TO THE USE OF LISTENING ASSESSMENT INSTRUMENTS

Charles V. Roberts
McNeese State University
Lake Charles, Louisiana 70609
(318 - 437 - 5039)

Presented to the Speech Communication Association Convention,
Denver, CO, November, 1985

BEST COPY AVAIL.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles Roberts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 505 120

ABSTRACT

A User's Response to the Use of Listening Assessment Instruments

This paper examines five well known listening measurement devices from the viewpoint of a "user" of listening assessment instruments. The strengths, weaknesses, procedural problems, and conceptualizations of each are considered. Applications of each are suggested and future needs in the area of listening measurement and research are discussed.

BEST COPY AVAILABLE

A User's Response to the Use of Listening Assessment Instruments

Many studies have verified the premier finding of Rankin (1926) that listening is the most frequently used mode of human communication. Indeed, almost every paper written concerning listening includes, early on, this insightful comment. At first blush, one wonders, then, why it is that we have to keep stating the obvious and feel compelled to support the assertion with footnotes citing Rankin, Nichols, etc. The answer very well might be that simply saying something does not mean it will be heard, and simply hearing something does not mean it will have any effect upon the subsequent behavior of the receiver. The receivers of our assertions are the members of our discipline. Despite our incessant pleading, our discipline's attention to listening does not mirror the voiced importance alluded to above.

Pedagogical consideration of listening seemed to dramatically increase during the 1950's and early 1960's, perhaps because of the availability of the Brown-Carlsen Listening Test (Brown & Carlsen, 1955). In the mid-1960's, a number of criticisms of listening tests, and indeed of the whole conceptualization of listening, surfaced. Perhaps as a result of these criticisms the number of published studies of listening declined. There is less research on listening published in our journals now than there was in the 1950's. This relative paucity of research is reflected in basic speech textbooks and, perhaps more critically, in listening texts. Two recent basic listening textbooks footnote as many studies done prior to 1960 as they do studies done after 1960 (Steil et al, 1983; Wolvin et al, 1982). Other scholarly works fare no better. In 1978, ERIC and SCA jointly issued Assessing Functional Communication (Larson et al, 1978) in which listening assessment was discussed. Only one of the references cited in the article was written within five years of the publication date of the article while nine were written prior to 1960.

There are indications that this decline in scholarly attention is ending. A new organization, the International Listening Association, is gaining strength. The business community is increasing its emphasis on listening training. Perhaps most striking of all, the number of listening assessment devices available to the communication researcher and/or teacher is increasing. The Brown-Carlsen test remains the most popular assessment device, however both the Watson-Barker Listening Test (Watson & Barker, 1984b) and the Kentucky Comprehensive Listening Test (Bostrom, 1983) are gaining popularity. The Communication Competency Assessment Instrument (Rubin, 1982) has been utilized at a number of locations, and the Learning Skills Inventory (Heun, Heun & Schnucker, 1977) approach to the assessment of communication skills has drastically changed the advising and teaching patterns of several universities.

All of these instruments have many positive attributes. Though all manifest some weaknesses, all function well when used for appropriate purposes. The remainder of this paper will focus on the utility of the various instruments and attempt to highlight problems that potential users should be aware of before selecting any one of the instruments.

BEST COPY AVAILABLE

Learning Skills Inventory

The Learning Skills Inventory, unlike other instruments, does not purport to measure how well a person *does* communicate. Rather, through a series of responses to 220 questions, a learning skills profile is developed. These self-evaluations provide insights concerning how well the respondents *think* they relate, symbolize, think, and remember. This self-report instrument also assesses the affect component of communication by asking questions concerning what the respondents "like to do." This is an important strength of the Learning Skills Inventory. McCrosky (1982) points out that assessment instruments that focus on skills typically address themselves to the question of whether or not a person *can* do something, and not on whether they *do* do something. While it is important that people learn the required competencies and skills, he cogently points out that the goal of instruction is "ultimate behavior." He suggests that failure to communicate competently may not always be related to lack of skills or competencies, "but rather may be the product of affective inhibition in people who are both competent and skilled"(p.6). Different people need different training. "Some need to develop skills. Others need to alter their communicative orientations and feelings. Accurate diagnosis should precede instruction. Confusing competence with performance and/or ignoring affect will lead to both inaccurate diagnoses and ineffective instruction" (McCrosky, 1982, p. 7). Knowledge and skill are not enough to predict effectiveness. A person's motivation to do so must be entered into the equation as well. This is precisely the problem that the Learning Skills Inventory attempts to solve. Respondents disclose what they perceive is their own relative development in each of four major skill areas, broken down into fifty-five sub-skill areas. High, middle, and low strength skills are indicated. The reliability of this test has been extensively tested and is acceptable. When placed in computers, the test is easy to take and can be easily scored, with the results quickly dispensed to the respondent.

Were users only provided this test of perceived strengths and weaknesses, the usefulness of this data would vary greatly, depending upon the perspicacity of the test administrator. Fortunately, such is not the case. Along with the instrument, the developers provide the user with a "guidebook" that helps the user interpret the results of the test and, most importantly, formulate adaptive learning strategies. The data allow for mapping individualized alternative pathways for learning. At colleges where this instrument is employed, students use the information to develop their own strategy for gathering information and obtaining meaning from that information. They and their advisors and teachers use this information when deciding on a program of skill enhancement. Thus this information allows the student to learn most effectively, while at the same time indicating skill areas that need work. The same instrument has been a help to teachers and business persons in assessing their own strengths and weaknesses and in developing strategies to make better use of the former and plan methods for dealing with the latter.

There are some limitations associated with this instrument. Knowing that they feel they are weak in listening might motivate students to take courses that depend primarily on textbook information. However, if no such option is available to the students, frustrations are increased. Institutional users of this instrument need to be committed to providing alternatives that answer these needs. Further, students who feel they are weak in listening may also be motivated to take listening courses. The lack of such courses at many universities may further increase the

BEST COPY AVAILABLE

students' frustration.

This instrument is not suited to the task of competency or skill assessment. As with most self-report instruments, it is highly fakeable. It can not be used to assess the changes in either competency or skill levels. Nor can it be used to judge the relative competency or skill of two or more people. While the creators of the test rightly state that the test has face validity (users are asked to respond to statements like "I am a good listener")for the purpose of discovering self-perceptions, no claim has been or can be made concerning its validity as a measure of listening competence or skill. It has been my experience that some very fine listeners (subjectively judged by a panel of one expert- myself) have judged themselves weak and many poor listeners (located using the same unreliable method) have assessed themselves as being strong in this area. This problem is not unique to this test. It may be that some people are poor judges of their own abilities, or it may be that their working definition of concepts like "good listening" differ markedly from my own. Finally, the test is administered in written form, and is thus dependent upon the ability of the user to read effectively. Respondents need to be able to read and comprehend the test, a presupposition that we can no longer automatically make concerning our college students. While the vocabulary level *should* prove no barrier to the typical college student, the current trends of lowering ACT scores at open enrollment institutions and of increasing numbers of students in remedial classes *should not* comfort the future user. Future versions may have to be administered orally. Such a procedure would allow attempts at explaining the various statements. This is an option of the Communication Competency Assessment Instrument and might be worthwhile if inter-user consistency is deemed desirable.

Communication Competency Assessment Instrument

The Communication Competency Assessment Instrument (CCAI) shares one characteristic with the Learning Skills Inventory in that both attempt to assess multiple skills within the same testing procedure, but there the commonalities cease. A growing number of institutions use the CCAI to assess a variety of communication skills of college students in order to determine if they have attained certain competency levels. Unlike the Learning Skills Inventory, users of the CCAI are asked to perform certain communication tasks. Listening skills are measured directly, without the use of "indirect paper/pencil instruments" and " in situations with which all students are familiar. Thus, the test was constructed around an educational context and provides measurements of how students communicate in classes, and with their professors and peers"(Rubin, 1982, p.xxi). Users are asked to attend to a videotaped presentation of a short lecture, much like one that would begin a college Listening course. To my knowledge, this is the only assessment device that uses videotape. Though studies have not indicated that this channel is significantly more effective than others (i.e. audio-tape), it does increase my estimation of the face validity of the test. I have few students who attend my words without also receiving my gestures and other nonverbal, visual messages. Some research has indicated that the dynamism and trustworthiness of the speaker are related to the long term retention of information (Roberts, 1980) and viewing a speaker delivering a speech gives clues about these two factors. Based on this evidence, one could argue both for and against having video-taped presentations. My argument would be for the affirmative side, though I would like to see a variety of presenters in this formate rather than just one. Other listening tests have utilized several presenters to

BEST COPY AVAILABLE

offset the possibility of speaker/message/receiver interaction effects. Varying the sex, age, race, ethnic characteristics, etc. of the senders is prudent if overall listening competence is to be assessed. Though the same caution should be exercised with regards to the message content, I like the fact that the content concerns listening. Even if the student lacks listening skills, the test itself becomes a "learning experience." The test's validity has been assessed " by having four of five communication faculty members agree on a blind placement of the stimulus questions into the correct competency areas (Rubin, 1982, p. xxii)". The creator of the test maintains that the " test is valid and the items are conceptually consistent" (Rubin, 1982, p. xxii).

Another unique characteristic of the CCAI is the reliance on the oral response mode rather than on the "indirect" paper and pencil responses of many other tests. Given the nature of the tasks students are asked to do, this is a significant strength. The problems of possibly measuring ineffective reading and writing skills are overcome. It does create some potential problems with rater reliability, but none so significant that they could not be overcome with training. However, students who are weak encoders could still demonstrate problems with the listening competencies even if they were not really deficient. Of particular concern are those students with severe communication apprehension. They might not have any problem listening and recalling the information and would test out as superior students given a written response mode, but would be found deficient if tested using the oral response mode. Given that the intended use of this instrument is to identify communication problem areas, this could lead to a false diagnosis. Since I have used the listening sections of the test in isolation, I have had to be especially careful in this regard. If users administer the total test they should be able to discern such problems during the encoding competency sections. The time element is a problem for those who are only interested in the listening competency and still wish to make sure that the individual is not a high apprehensive. I am somewhat distressed by those who are calling for a written multiple-choice response mode for the listening section of the CCAI (Rubin & Shepherd, 1985). While this would make large-scale testing situations practical, it would alter the nature of the task being asked of the students and, I believe, take away several of the truly unique and positive characteristics of the CCAI.

The CCAI is designed to assess understanding and short-term recall as well as the user's ability to "recognize a fact in a class lecture or report"(Rubin, 1982). Testing *recall* rather than *recognition* is one of the unique characteristics of this instrument. Most tests of listening ability are set up to test recognition rather than recall. The two skills are not the same. Some theorists hold that recognition retrieves information from a different memory store (semantic store) than does recall (episodic store) (Cofer, 1975). This would not be a telling difference if, indeed, the test did assess *short-term memory* as is maintained (Rubin, 1982; Rubin & Shepherd, 1985). Both semantic and episodic memory stores are conceptualized as long-term memory. However, after administering the test many times, it is clear to this user that the instrument is tapping long-term memory. The taped lecture runs for more than seven minutes and the questioning procedure clearly extends the period between stimulus presentation and recall beyond the usually accepted duration of short-term memory (Cofer, 1975; Bostrom, 1985). This is not a severe conceptual flaw. I greeted this realization as a serendipitous happening. While certain situations do require effective short-term memory, many more require long-term retention. I find a measure of long-term memory recall ability much more useful than one that would demonstrate short-term recall or recognition ability. This is

BEST COPY AVAILABL

especially so for the listening context that this instrument focuses on - the classroom. That the CCAI does measure long-term memory rather than short-term memory is demonstrated by the research of Rubin and Shepherd (1985). The correlations with the "lecture listening" sections of several other tests and the lower and/or lack of correlation with the short-term sections are support for my view.

Finally, I have not found the CCAI as suited for research as some of the other listening tests. The time factor, of course, weights heavily against its use. Further, its original intent was "to provide a measure of students' readiness for upper-level college coursework" (Rubin, 1982, p. iv). The test is not designed to be used as a measure of competence for any one skill. This is just as well since some of the scoring seems to require a high level of rater expertise to insure reliability. For example, in scoring "Competency #8," the rater would give the respondent a "three" if the student answered the question correctly, but did not "sound certain," a score of "four" if the student gave the same answer, but this time ***with certainty***, and a top score of "five" if and only if the student "sounds certain that she/he is identifying a fact and gives a brief reason why (referring to the study on which it is based)" (Rubin, 1982, p.8). The question the student responds to seems to be a closed one. Hence students who follow instructions carefully are not given as much credit as those who go beyond the limits of the question. Raters not only must read nonverbals to ascertain "certainty," but also keep their own nonverbals under control so as to allow for the student to freely add information without being prompted. When I use the test, I find it useful to ask students what they base their opinion on if they do not indicate this to me freely. Many respond with the full and complete answer the highest score requires.

Brown-Carlsen Listening Comprehension Test

The Brown-Carlsen Listening Comprehension Test was the first standardized listening test - though there were a few that predated it that were not standardized. Since it was first administered over 35 years ago, this test "has probably gone through more trials, revisions and refinements than most tests" (Brown, 1985).

Brown (1985) has demonstrated that this test measured a unique skill that does not correlate highly with reading ability(as measured by the Nelson-Denny Reading Test), intelligence (as measured by the ACE), or scholastic achievement (as indicated by grade point average and class rankings). Further efforts were made to establish validity "(1) by definition, (2) by subtest interrelationships and (3) by subtest consistency" (Brown, 1985, p.2). All showed support for this operationalization of listening ability. Other tests indicated that the measure was both reliable and appropriate for high school and college students.

One method for judging the "worth" of a measurement tool such as a listening test is to see how much it has been used. By that criterion alone, one would have to grant great value to the Brown-Carlsen Listening Comprehension Test. For more than thirty years, the majority of researchers interested in studying listening have utilized this test. Of course, usage alone has nothing to do with the value of the findings gained through the administration of this instrument. As Watson and Barker (1984a) point out, results of listening studies "are of limited value unless the instruments are reliable and valid measures of listening comprehension" (p. 187). As is, and should be, the case with almost every major measurement device, the reliability and validity of the instrument has been scrutinized. In the 1960's a number of critics looked

BEST COPY AVAILABLE

carefully at the the leading listening measurement instruments (See, for example, Becker, 1963; Petrie, 1964; and Kelly, 1963, 1967).

Perhaps the most important criticism of listening tests was leveled by Charles Kelly. Among his findings was that the Brown-Carlsen test and the STEP test correlated more highly with a test of intelligence than they did with each other. When judging the adequacy of this line of argument, it should be remembered that Brown presents evidence that his instrument does not correlate with measures of general intelligence. Nonetheless, modern critics of listening tests often cite Kelly's finding, interpreting it as an indictment of existing listening measurement techniques and a rationale for the inditing of "more valid" tests of listening ability. I believe that this interpretation is not totally accurate. For me, the major thrust of his study was to differentiate between ***listening performance and listening ability.*** To develop his argument, Kelly (1963) cites Stromer(1952), Hackett (1955), and others as he builds his research case. None of these sources question the internal validity of listening tests so much as they do the generalizability of listening research. While Kelly does criticize both the internal and external validity of listening instruments, his attack on the latter is the more telling. He forcefully argues that "we have a massive body of information about listening behavior of subjects who knew they were going to be tested. This is important information dealing with one type of listening activity. But we have done almost nothing to find out about performance across the general range of situations from panic to boredom" (1967, p. 464). Kelly concludes that "our traditional procedures for testing listening are sterile, as customarily used, and that currently published listening tests are not valid measures of a unique skill such as has been posited in much of the literature on listening" (1967, p. 455).

It is interesting to note that though many have accepted his criticism as valid, the majority of the creators of modern listening assessment instruments continue to test for listening competency and skill in situations where anticipatory set would be assumed to be functioning. Under such situations, Kelly's finding that mental ability and listening ability are correlated is not unexpected.

> It is likely that, as a subject's motivation to listen increases, the influence of mental ability upon his comprehension will also increase. In other words, when a listener's attention is maximal (as when taking a test), he probably makes full use of his mental ability to comprehend what is being presented, and his personality traits, past listening habits, etc. are relatively less important (Kelly,1967, p.464).

It has been suggested that, rather than follow Kelly's suggestion and test for listening competency and skill outside of the "motivating environment," that we continue to control for motivation by alerting the test takers thus equalizing motivation.

> Our judgments of either competence or skill must be based on observations of overt behavior. Such judgments should be based on carefully controlled situations in which the person to be judged is aware that his/her competence/skill is to be observed and evaluated, and in circumstances in which the person is motivated to be perceived as competent or skilled. The typical classroom situation may provide such a setting. Under such circumstances it is possible to determine whether the person *can* engage in the competent or skilled behavior. It is not possible,

BEST COPY AVAILABLE

> however, to judge whether the person *will* engage in such behavior in later life. Both competence and skill are abilities which are mediated by motivations in everyday life and cannot be expected to be universally manifested in behavior under all circumstances (McCrosky, 1982, p. 7).

The internal validity of the Brown-Carlsen test is another question. Even if limited to the motivating atmosphere of the testing situation, information concerning a subjects ability to listen would prove worthwhile. Whether the Brown-Carlsen instrument does tap listening ability is an important question. Many have suggested that listening is a complex process, rather than a simple, unidimensional skill. This view is not disputed by the developers of the Brown-Carlsen test. Five subtests of the instrument measure the respondent's ability to follow directions, recognize transitions, recognize word meanings, and recall information immediately after it is presented and at a delayed time (lecture comprehension). But is this the "complex of behaviors" that today's theorists call listening ? The answer to this question is an emphatic **YES** followed closely by a resounding **NO!** Indeed, there does not appear to be any clear cut concensus concerning the "*correct* " definition of that elusive concept. As Weaver points out "standardized tests...were built to measure skills their authors decided were the critical subskills, and no two tests measure the same behaviors"(1976, p. 17). If any doubt this assertion, they need only attend to the reasoning of modern theorists concerning the conceptualization of listening (See, for example, Bostrom, 1985; Watson & Barker, 1985).

Differences in conceptualization and operationalization of variables abound in the literature. Such diversity is neither a surprise, nor a "curse." It allows for the emergence of the most robust theoretical explanation. As a result of the current "definitional jousting" several newer listening measurement devices have been created to take the place of the Brown-Carlsen Listening Comprehension Test. Partial justification for these instruments seems to be the questionable validity of the Brown-Carlsen test. Given that other instruments were believed unsuitable for certain research efforts, and rather than cease doing research, at least two new instruments were created within the past five years to measure the complex concept that we call listening.

Watson-Barker Listening Test

One such instrument is the Watson-Barker Listening Test. This test was developed in 1982 in an attempt to create a standardized listening test that would be oriented primarily toward adults and mature college level audiences. Its "face validity was assessed by using a panel of listening experts to judge the validity of each item"(Watson & Barker, 1984b, p. 1). Additional support for the validity of the instrument has been generated by Rubin and Shepherd (1985) and by Applegate and Campbell (1985). Both studies link this instrument with other listening measurement tests. While such experiments will help to establish the efficacy of comparing data of the various tests, they provide only a tautological validation of the instruments. If all tests are highly correlated and if any one test is valid, the the validity claims of all tests can be accepted. If no check of validity other than that of "face validity" is performed, all such claims should be held in abeyance until the concept of "listening" is agreed upon substantively by listening theorists. Roberts (1985) has presented some evidence that the Watson-Barker Listening Test does correlate in predictable ways to the Receiver Apprehension Test. Given the

BEST COPY AVAILABLE

amount of criticism directed at the validity of the Brown-Carlsen instrument, more studies need to be done before the majority of users accept the validity claims of this, or any newer conceptualization of listening measurement.

Researchers can and do operationalize terms as they wish. If one accepts the operationalization, then there is no difficulty concerning utility. The Watson-Barker instrument conceptualizes listening as a combination of receivers ability to evaluate message content and emotional meaning in messages, understand meaning in conversations, understand and retain lecture information, and follow instructions and directions (Watson & Barker, 1985). I have used this instrument for several reasons. I feel that the sub-skills it taps are more in line with my research interests than some other instruments. It is easier and quicker to administer than other competing tests. Though I would prefer to tap recall as well as recognition, I am willing to sacrifice this factor when the number of subjects I need to test is large. I have had the experience of coding the responses of over two hundred subjects on one immediate and three delayed recall tests of material presented in a naturalistic setting. I would have to be very interested in a project before I agreed to a similar operationalization in the future (or have several graduate students under my thumb). The CCAI does not allow for the degree of precision necessary for much of my statistical analyses, or it would provide an option that I might consider.

I like the fact that most of the Watson-Barker stimulus material is capable of being generated in a non-laboratory setting. This helps defend its generalizability. A fitting test of the instrument would be to have subjects respond to questions that would mirror the content of this test under conditions of "nonawareness of the intent to test," and then to compare those data with scores generated "under testing conditions." The enormity of the effort required prompts me to unselfishly offer the task to anyone else who might care to have it. Watson and Barker are working on a video-taped version of their instrument. This should increase the "reality" of the stimulus material.

There is some problem with "cheating" on the exam. Respondents can "look ahead" to the questions, gaining hints as to future questions. Though this is not a necessary "flaw" in the test and has more to do with its administration, future users should be aware of this and control for it. Additionally, the version of the test that I am most familiar with does not contain a "distraction" segment (though more current versions with distractions are available). Nonetheless at the end of each page it became necessary to stop the tape, allowing people to turn the pages of the test booklet without having the disturbing sound of papers bother the reception of other test takers.

While I have used the Watson-Barker test in research situations, I have found it more useful as a "consciousness raiser" in my basic speech classes and in various seminars offered for industry. It has proven a quick method for demonstrating the need of further listening training for both constituencies. The self-scoring answer sheet is especially welcome. While using it in one seminar I did uncover a possible problem with some of the stimulus material. It seems that people who have experience with instructions similar to those presented on the tape (i.e. -copy machines) have fewer problems with those areas even though their ability to answer other questions in the same section is inadequate. Future users may wish to evaluate this "history" component in their own respondents.

BEST COPY AVAILABLE

Kentucky Comprehensive Listening Test

The Kentucky Comprehensive Listening Test (KCLT), like the Watson-Barker instrument was created just a few years ago and is oriented towards adults. Like the Watson-Barker test, the KCLT has served to spur research activity these past several years. Again, like the Watson-Barker test, the KCLT is claimed by its developer to measure a variety of different sub-skills. The similarities continue in that both are mediated by other communication skills, and both are delivered via audio tape.

In terms of usage, I have found the KCLT less amenable, in its full form, for my research and "consciousness-raising" purposes because of its length. The lecture material was considered too drawn out and uninteresting by several groups. Possibly in response to experiences such as mine, Bostrom (1984) has suggested that, for research projects, users dispense with the lecture section. My respondents also found the short-term retention tasks unrealistic, and were prone to "cheat" by looking at the various possible answers before hand and developing methods for "remembering" the correct answer.

In several instances there has been an interesting response to the "distraction" section of the test. It may be because of the naivete of my subjects, but for whatever reason, they were especially interested in the *very* distracting conversation that is used to disrupt the listener's concentration. Those that demonstrated intense interest in the distraction and could remember, with almost total recall, the conversation, also scored rather well on that section. If this audio-taped version is ever converted into a video-tape, many will want to view the young lady in the role of the "distracter."

Of crucial consideration for potential users are the **dissimilarities** between the Watson-Barker test and the KCLT. As indicated previously, users choose measurement instruments partially on the basis of the underlying conceptual process that is operationalized by the test. For Bostrom that underlying process is, of course, listening. However, unlike Watson and Barker, he discounts several possible sub-processes that they suggest partially make up listening, and concentrates on short-term memory, short-term memory with rehearsal, "lecture listening" (long-term memory), selective listening, and interpretive listening (Bostrom, 1985). He discounts the the sub-process of comprehension reasoning that "common sense tells us that we can listen without full understanding --in fact, often the question of understanding is irrelevant" (1985, p.4). As a user I am aware that people can listen without full understanding, indeed episodic long-term memory is conceptualized to be made up entirely of "nonverbal, nonsymbolic" information. However, I also believe that we store information in semantic memory which, as conceptualized of some, must include the concept of comprehension. Limiting retrieval tasks to "recognition" would make "comprehension" less necessary, but expanding the test to cover the full spectrum of memory tasks would demand its inclusion.

Bostrom likewise has a very powerful argument against including the affect dimension in any consideration of listening. He points out that we can not specify any observable behavior that would allow an observer to ascertain when someone else is listening effectively. "The real

BEST COPY AVAILABLE

poverty of an attitudinal approach is dramatized when one attempts to codify it into some concrete behaviors which will result in 'good' listening. Beyond acting interested and not looking at one's watch while the other is talking, there is little to do to convince others that one is <u>really</u> listening"(1985, p. 6). From a user's standpoint such an argument is not persuasive. We often rely on self-report instruments. If we accept his viewpoint, what are we to do when we attempt to assess attitude change? Evidence seems to indicate that there are no necessarily conclusive behavioral indications that people believe or feel any particular way. Further, is it so important for the listening teacher to instruct his charges on how to *seem* like they are good listeners, or is it more important for the teacher to help her pupils *be* good listeners? Tests like the <u>Learning Skills Inventory</u> attempt to assess affect. Perhaps elements of the two together might be combined to yield a richer assessment of both listening ability and performance - an outcome seemingly called for by Kelly twenty years ago.

While I do disagree with some of Bostrom's reasoning concerning the conceptualization of listening, I strongly agree with his feeling that people listen differently in different situations. I feel that listening is self-directed. People listen as they wish to, either consciously or subconsciously. These desires are under their control - listening effectively is the listener's responsibility. Speakers can get people to listen, but only by persuading them that it will be useful for them to listen. For the sender the goal is to "get the receivers interested in listening." It is my understanding that Bostrom's functional approach would allow for this stance.

He points out many advantages of this approach when he states that first

> it provides a comprehensive theoretical model based on fairly well-known memory functions; and fits well into private-public models of communicative behavior. Second, it provides a comprehensive answer to the problems originally raised by Kelly and ignored by researchers since the middle 1960's. Third, it points to new directions in listening research, an area substantially ignored by researchers for a number of years (Bostrom, 1985, p. 16).

Notwithstanding our different placing of emphasis concerning Kelly's critique of listening measurement, his point is well taken. His analogy to the various types of listening is intriguing as well, as he wonders " ... if ordinary 'sending' behavior is of many types, should not the receiving activity associated with it also vary?"(1985, p.8). A word of caution may be required with regards to pursuing this line of investigation. Cronhkite (1974) suggested that research be undertaken to investigate "variables that influence the audience's ability to reliably evaluate messages," and to "turn our existing speaker-oriented research upside down to discover implications for critical listening" (pp. 81-82). Sprague (1974), too, called for the translation of "speaker-oriented, control-oriented theories and research findings into receiver-oriented, choice expending implications"(p. 83). While the pleas of such scholars for the creation of listening theories is persuasive, the successful translation of sender theories into receiver theories has not yet happened. It may be that it never will. Crucial to the success of such a venture is the implied linkage between encoding and decoding. If we are to flip these theories over so that they address themselves to effective listening rather than effective speaking, should we not first ascertain if there is such a connection? So far such connections have been suggested by many, accepted axiomatically by some, and substantiated by no published research that I have discovered. The fulcrum that would allow us this Atlas-like task remains

BEST COPY AVAIL

elusive. We have yet to ascertain if listening and speaking "mirror" each other or "shadow" one another. Does one process reverse its opposite, repeat it, or are they totally different from each other? If they are reversible, then we can "turn our... research upside down." But if the latter is the case, data derived theories of speaking "only" need be generalized "right-side up" to apply to listening situations.

Currently a definition battle is being waged, albeit a refined, scholarly one, concerning the conceptualization of listening. Bostrom presents ample evidence of the validity of his measure, as do all of the creators of listening measurement instruments. One such claim is that "each of the scales represents an actual instance of the performance of the skill in question" (Bostrom, 1984, p.2). With claims such as this, he and others seem to be attempting to avoid a full scale war by begging the question that while the definition of listening has not been agreed upon, the various sub-skills that the various tests measure have been universally accepted. If the "whole" has not been agreed upon, then the "parts" that make up the totality of that "whole" are no surer. While all, save the measurers of affect, would include "retention" within the conceptual framework of listening, what aspects of retention are important and how they are to be tapped is not so certain.

Conclusions of a User

It seems certain, to this user, that the debate as to which instrument is the most appropriate measure of listening competency, skill, and/or performance will continue. It seems just as certain that the definition battle outlined above will be waged concurrently. One instrument can not win acceptance without its corresponding underlying conceptual definition being agreed to by the majority of users. Perhaps no ***ONE INSTRUMENT*** will be found to be acceptable for all situations. In any case, we should not be upset by such jousting. Nor should we stand dispassionately aside as the "contest for acceptance" goes on. We users should enter in to the arena as "interested third parties" acting in the role, perhaps, of "devil's advocates." Subjecting these measurement instruments to the fire of our scrutiny will result in much more sound listening test(s). The need for such a tool is evident. Without it, we can not hope to develop effective methods of listening instruction. Given the rather sketchy evidence available, it is difficult to argue with Erway's (1972) contention that gains from listening instruction are not maintained over time. Most listening research studies are "quick and dirty." Few longitudinal studies have been done. The generalizability of most studies is severely limited by the nature of the subject population drawn upon. By far the most prevalent educational level in listening research is the elementary school level. Fewer studies have been carried out at the secondary level, and fewer still have been completed using college-age subjects. This inverse relationship between the amount of studies and the age of subjects seems to mirror the relationship between age and potential for listening improvement that some researchers have alluded to in their articles (Evans, 1960; Everts, 1962; Lieb, 1965).

A close reading of most listening texts reveals that there is little reason to support the contention that we currently are teaching listening effectively. For such support we continue to have to fall back upon the subjective judgments of other teachers of listening. Erway (1972) has suggested, "the most impressive evidence comes not from research but from the prejudiced reports of students who have experienced instruction and from the observation of instructors" (p.23). This "evidence" must be considered especially suspect in light of the finding that people

BEST COPY AVAILABLE

tend to think more highly of themselves as listeners than test scores indicate and that they are less able to discriminate between good and poor listening than they are between good and poor speaking (Stark, 1956). While it can be argued persuasively that we should teach listening at all educational levels, the only well documented listening finding is that listening is not being taught in most academic institutions.

Implementing longitudinal investigations that would document effective methods for teaching listening would help to reverse this tendency towards lip service. If scarcity does increase the value of a commodity, the results of such studies done in the classroom situation would prove very worthwhile. Prior to 1970, only fifteen empirical studies investigated pedagogical phenomena by first teaching teachers to behave in some particular way, then observing them to make sure they did behave in that way, and, finally, testing their students to note changes (Sprague, 1974). As noted previously, there are pronounced problems in generalizing laboratory research to the classroom and beyond. What is lost in terms of ability to control and limit experimental artifacts would be made up for in terms of the vigor and power of the generalizability of the resultant data.

Until the majority of our field, who are interested in listening research, agree upon a definition of listening and instruments to tap that conceptualization, we users will not be able to proceed with our full attention to develop such experimental paradigms. Until such experiments are conducted, teachers interested in increasing listening skills can do no better than rely on the unsubstantiated platitudes that currently make up the bulk of our listening instruction. We will continue to tell our students to "Withhold evaluation of the message until the speaker is finished" (Barker, 1984, p. 55) and hope they don't ask us too many questions about the research that indicates that that is appropriate behavior. There is no research documentation that would support such imperatives. One even could argue that such a course of action is inefficient since it causes you to listen to unimportant as well as senseless drivel. Further, even if that inefficiency were shown to be necessary and/or useful, no pedagogical direction is available that would allow a teacher to help students carry out that directive. How does one "withhold evaluation" on the attitudinal level? Does the evaluation only matter if done on the "conscious" level? Does it matter if people do evaluate a speaker, if they still continue to listen to him?

We need to develop measures that are valid measures of listening, regardless of where and under what circumstances that activity takes place. Perhaps several instruments will be needed to cover all of the important contexts we wish to tap into. Expediency necessitates that we then undertake investigations to ascertain how we can best facilitate more effective listening. It may well be that our listening texts have more substance than alluded to above. If research reveals that there are founts of knowledge and potent developers of skills already extant, more weight can be applied in the effort to wedge in listening instruction in our already crowded curricula. If none of our current teaching imperatives are supported, future research directions will be more clear and the weight of unsubstantiated dogma will no longer have to be borne by listening instructors. Whichever the case, we need to go forward. "As long as we lack such research we shall be bound to myths and superstitions which are interesting subject matter for our methods courses, but which have little relevance for the real world "(Sprague, 1974).

BEST COPY AVAILABLE

REFERENCES

Applegate, J. S. & J.K. Campbell. A correlation analysis of overall and sub-test scores between the Watson-Barker and the Kentucky Comprehensive Listening Tests, paper presented at the International Listening Association Convention, March, 1985.

Barker, L. Communication. Englewood Cliffs, N.J.:Prentice Hall, Inc., 1984.

Becker, S. Relationships among the communicative skills, Central States Speech Journal, 14 (1963), 258-264.

Bostrom, R. Conceptual approaches to measuring listening behavior, paper presented at the Speech Communication Association Convention, November, 1985.

Bostrom, R. The Kentucky Comprehensive Listening Test. Lexington: Kentucky Listening Research Center, 1983.

Bostrom, R. Research Update. Lexington: Kentucky Listening Research Center, 1984.

Brown, J. The construction of a diagnostic test of listening comprehension, Journal of Experimental Education, 18 (1949), 139-146.

Brown, J. The Brown-Carlsen listening comprehension test, paper presented at the Speech Communication Association Convention, November, 1985.

Brown, J. and G. Carlsen. Brown-Carlsen Listening Comprehension Test, Manual of Directions, Yonkers-on-the-Hudson, New York: World Book Company, 1955.

Cofer, C. The Structure of Human Memory, San Franscisco: Freeman & Co., 1975.

Cronkhite, G. Rhetoric, Communication and psychoepistemology: An academic exercise for the Sophistic devil, in Rhetoric: A Tradition in Transition (W. Fisher, ed.), East Lansing, Michigan: Michigan State University Press, 1974.

Erway, E. A. Listening: The second speaker, Speech Journal, 10 (1972), 22-27.

Evans, A. Listening related to speaking in the first grade, Master's Thesis, Atlanta University, 1960.

Everts, E. An investigation of the structure of children's oral language compared with silent reading, oral reading, and listening comprehension, unpublished Doctoral Dissertation, Indiana University, 1962.

Hackett, H. A null hypothesis: There is not enough evidence. Education, 75 (1955), 349-351.

BEST COPY AVAILABLE

Heun, R., L. Heun, & R. Schnucker. Learning Skills Inventory, Kirksville, MO: Learning/Instruction Facilitators, 1977.

Kelly, C. Listening: Complex of activities - and a unitary skill? Speech Monographs, 34 (1967), 455-465.

Kelly, C. Mental ability and personality factors in listening, Quarterly Journal of Speech, 49 (1963), 152-156.

Larson, C., P. Backlund, M. Redmond, & A. Barbour. Assessing Functional Communication, published jointly by Urbana, Illinois: ERIC Clearinghouse on Reading and Communication Skills and Falls Church, Virginia: Speech Communication Association, 1978.

Lieb, B. The relationship between some aspects of communicative speaking and communicative listening in freshmen men and women, unpublished Master's Thesis, Pennsylvania State University, 1960.

McCroskey, J. Communication competence and performance: A research and pedagogical perspective, Communication Education, 31 (1982), 1-8.

Nichols, R. Factors in listening comprehension, Speech Monographs, 15 (1948), 154-163.

Petrie, C. What we don't know about listening, Journal of Communication, 14 (1964), 248-252.

Rankin, P. T. The measurement of the ability to understand spoken language, unpublished Dissertation, University of Michigan, 1926.

Roberts, C. The effects of ethos on the retention and reminiscence of information, unpublished Dissertation, Temple University, 1980.

Roberts, C. Preliminary research employing the Watson-Barker Listening Test: A validation of the instrument, paper presented at the International Listening Association Convention, March, 1985.

Rubin, R. The Communication Competency Assessment Instrument, Annandale, VA: Speech Communication Association, 1982.

Rubin, R. & P. Shepherd. Listening assessment via the Communication Competency Assessment Instrument, paper presented at the Speech Communication Association Convention, November, 1985.

Sprague, J. Cognitive skills and teaching for mastery, in New Horizons for Teacher Education in Speech Communication (Newcombe, P. J. and R. R. Allen, eds.), Skokie, Illinois: National Textbook Company, 1974.

BEST COPY AVAILABLE

Stark, J. An investigation of the relationship of vocal and communicative aspects of speech competency with listening competency, unpublished Doctoral Dissertation, New York University, 1956.

Steil, L., L. Barker & K. Watson. Effective Listening: Key to Your Success, Reading, Massachusetts: Addison-Wesley Publishing Company, 1983.

Stromer, W. Listening - How? English Journal, 41 (1952), 318-319.

Watson, K. & L. Barker. Listening Assessment: The Watson-Barker Listening Test, paper presented at the Speech Communication Association Convention, November, 1985.

Watson, K. & L. Barker. Listening behavior: Definition and measurement, in Communication Yearbook Eight(R.N. Bostrom, ed.), Beverly Hills, CA: Sage, 1984a.

Watson, K. & L. Barker. Watson-Barker Listening Test. New Orleans: Spectra, Inc., 1984b.

Wolvin, A. & C. G. Coakley. Listening, Dubuque, Iowa: William C. Brown Company Publishers, 1982.

BEST COPY AVAILABLE